Design Space Exploration of Turbulent Multiphase Flows Using Machine Learning-Based Surrogate Model

Himakar Ganti, Manu Kamin and Prashant Khare *

Department of Aerospace Engineering, University of Cincinnati, Cincinnati, OH 45221-0070, USA; gantihr@mail.uc.edu (H.G.); kaminms@mail.uc.edu (M.K.)
* Correspondence: Prashant.Khare@uc.edu

check for updates

Abstract: This study focuses on establishing a surrogate model based on machine learning techniques to predict the time-averaged spatially distributed behaviors of vaporizing liquid jets in turbulent air crossflow for momentum flux ratios between 5 and 120. This surrogate model extends a previously developed Gaussian-process-based framework applicable to laminar flows to accommodate turbulent flows and demonstrates that in addition to detailed fields of primitive variables, second-order turbulence statistics can also be predicted using machine learning techniques. The framework proceeds in 3 steps—(1) design of experiment studies to identify training points and conducting high-fidelity calculations to build the training dataset; (2) Gaussian process regression (supervised training) for the range of operating conditions under consideration for gaseous and dispersed phase quantities; and (3) error quantification of the surrogate model by comparing the machine learning predictions with the truth model for test conditions (i.e., conditions not used for training). The framework was trained using data generated by high-fidelity large eddy simulation (LES)-based calculations (also referred to as the truth model), which solves the complete set of conservation equations for mass, momentum, energy, and species in an Eulerian reference frame, coupled with a Lagrangian solver that tracks the dispersed phase. Simulations were conducted for the range of momentum flux ratios between 5 and 120 for liquid water injected into crossflowing air at a pressure of 1 atm and temperature of 600 K. Results from the machine-learned surrogate model, also called emulations, were compared with the truth model under testing conditions identified by momentum flux ratios of 7 and 40. L_1 errors for time-averaged field quantities, including velocity magnitudes, pressure, temperature, vapor fraction of the evaporated liquid, and turbulent kinetic energy in the gas phase, and spray penetration and Sauter mean diameters in the dispersed phase are reported. Speedup of 65 was achieved with this emulator when compared against LES simulation of the same test conditions with errors for all quantities below 14%, thus demonstrating the potential benefits of using machine learning techniques for design space exploration of devices that are based on turbulent multiphase fluid flows. This is the first effort of its kind in the literature that demonstrates the application of machine learning techniques on turbulent multiphase flows.

Keywords: machine learning; large eddy simulation (LES); turbulent multiphase flows; gaussian processes

1. Introduction

Turbulent multiphase flows are ubiquitous and play an important role in numerous engineering devices including liquid-fueled propulsion systems, such as diesel, gas-turbine, afterburner, and ramjet engines. In particular, in afterburner and ramjet engines, the liquid fuel is often injected transversely into a high-temperature turbulent crossflowing air; the interaction between air and the liquid fuel

leads to atomization, vaporization, fuel/air mixing, and subsequent flame development. The flame is generally stabilized by a bluff body that is placed in the chamber with its axis perpendicular to the flow direction. Burned gases are entrained in the wake downstream of the bluff body, providing a continuous ignition source, thus stabilizing the flame [1]. This canonical configuration, where a liquid jet is injected perpendicular to an air stream is often referred to as liquid jet in crossflow (LJCF). It consists of rich fluid dynamic features, both in the gaseous and the liquid phases, that have been extensively studied in the literature both theoretically [2–6] and experimentally [7–12]. One of the major challenges in investigating this phenomenon is the presence of a wide range of length and time scales, which includes turbulence, chemistry, and acoustic scales in the gaseous phase, and scales related to hydrodynamic instabilities and breakup in the liquid phase. While significant progress has been made to enhance the understanding of such flows especially due to the development of advanced diagnostic techniques [7,13] and tremendous improvement in hardware and software technologies to simulate such flows [10,14–16], the disparate range of scales associated with the physical processes in this configuration renders the state-of-the-art techniques prohibitively expensive to conduct parametric studies under conditions that span the entire design space. Therefore, in this effort, we extend our previously developed machine learning (ML) framework that was applied to laminar multiphase flows in our previous paper [17] to turbulent multiphase flows. This ML-based surrogate model will be trained using data generated using large eddy simulations (LES) (also referred to as truth data from here on). Specifically, we will demonstrate the accuracy of the ML framework in vaporizing the liquid jet in a turbulent crossflow. Once trained, the ML algorithm will be used to predict the important flow variables including velocities, temperature, vaporized fuel fraction, spray penetration depth, droplet statistics, and the turbulent kinetic energy under testing conditions, that is, conditions not used for training. The error will be quantified by comparing these predictions with truth data.

Several researchers have developed empirical correlations to predict spray characteristics based on non-dimensional quantities, such as the momentum flux ratio, the Weber, and Reynolds numbers. For example, based on experimental observations, Wu et al. [16,18] developed correlations to predict the liquid column penetration over a range of momentum flux ratios and injector diameters. In another study, Amighi and Ashgriz [19] performed experiments on water injection in air crossflow for a range of momentum flux ratios under elevated temperature and pressure conditions. In this study, based on measurements of droplet sizes, trends were developed as a function of liquid and air velocities and ambient pressures. Many other researchers have developed such correlations to provide global estimates of droplet sizes and liquid jet penetration depths [11,20–22].

While estimates of spray characteristics are important, predictions of detailed spatiotemporal spray and gaseous flowfields will tremendously improve the efficiency and substantially reduce costs during the design process. In this context, in recent years, because of the availability of large datasets produced using high-resolution experiments and DNS/LES enabled by ever-increasing computational power, several research groups have developed surrogate models to predict fluid dynamic flows based on machine learning techniques [17,23–31]. A notable work is due to Ganti and Khare [17], who recently developed a Gaussian process-based machine learning (GPML) framework capable of predicting the spatiotemporal evolution of multiphase flows and demonstrated its accuracy and the associated speedup on two canonical laminar flow configurations—flow over a circular cylinder and injection and atomization of diesel fuel in a quiescent chamber. This was the first study of its kind to predict the entire spatiotemporal dynamics of liquid jet injection and atomization, including the initiation and growth of interfacial hydrodynamic instabilities, deformation, and subsequent primary and secondary breakup phenomena. In this manuscript, we will extend the applicability of this framework to turbulent flows, and show that in addition to primitive flow variables, second-order turbulent statistics can also be predicted using GPML. To demonstrate this, we will develop a GPML-based surrogate model to predict a vaporizing liquid jet injected perpendicular to a turbulent air crossflow with a triangular bluff body placed downstream of the injection location. The training and testing datasets will be developed using an Eulerian–Lagrangian large eddy simulation framework [32,33]. While the LES calculations will

be conducted in 3D, to showcase the capability of the GPML algorithm, we will only apply it on the mid-plane of the geometry once the calculation reaches a statistically stationary state. The operating conditions will be identified by the momentum flux ratio; all other relevant non-dimensional numbers will be kept constant.

This paper is organized into six sections. The next section provides a general description of the Gaussian process machine learning (GPML) technique, followed by the framework that we have established to predict turbulent vaporizing LJCF physics using GPML. Theoretical formulations underlying the LES framework that is used to generate the training and testing databases, and the quantification of error metrics, will be described in the subsequent two sections. Results in Section 5 are organized into three subsections—the first subsection presents the experimental validation of the LES framework and the selection strategy of the training and testing operating conditions. Predictions of the machine-learned algorithm and their comparison with LES data under testing conditions to establish the accuracy and speedup afforded by the surrogate model will be discussed next. The manuscript ends with major conclusions that can be drawn from the study.

2. Gaussian Process-Based Framework for Turbulent Multiphase Flows

In the next two subsections, we will first briefly describe the fundamentals of the Gaussian process regression technique and then provide the details of the machine learning framework that we developed for this study. For a detailed treatment on the development of the emulation framework, the Gaussian process machine learning algorithm, and its application to multiphase flows, the reader is referred to Ganti and Khare [17].

2.1. Gaussian Process Regression for Fluid Flows

The approach used to conduct Gaussian process regression (GPR) is similar to the approach of [34–36] and only a brief description is given. GPR does not give a functional form for mapping inputs to outputs, but it predicts the output y_* at an input x_*, along with the associated uncertainties, given n observations. The training data were assumed to have a Gaussian distribution with a zero mean and a covariance that can be defined by a kernel function, $k(x, x')$. Hence, the conditional probability $p(y_*|\mathbf{y})$ and the possible value of y_* at x_* has a Gaussian distribution. The covariance functions of the observations and the input point of interest x_*, in matrix form are given by the equation:

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{bmatrix}$$

$$K_* = \begin{bmatrix} k(x_*, x_1) & k(x_*, x_2) & \dots & k(x_*, x_n) \end{bmatrix} \text{ and } K_{**} = k(x_*, x_*) \tag{1}$$

The output y_* at the point x_* is obtained by estimating the likelihood of y_* given by the following Gaussian distribution:

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}\right) \tag{2}$$

The conditional probability is:

$$y_*|\mathbf{y} \sim N\left(K_* K^{-1} \mathbf{y}, K_{**} - K_* K^{-1} K_*^T\right) \tag{3}$$

with the mean and variance as:

$$\bar{y}_* = K_* K^{-1} \mathbf{y}$$
$$\text{var}(y_*) = K_{**} - K_* K^{-1} K_*^T \tag{4}$$

Kernel functions have hyperparameters associated with them, and Bayes theorem can be applied to obtain an optimum set of values for them by maximizing $\log p(\mathbf{y}|\mathbf{x}, \theta)$, given as:

$$\log p(\mathbf{y}|\mathbf{x}, \theta) = -\frac{1}{2}\mathbf{y}^T K^{-1}\mathbf{y} - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi \tag{5}$$

The output values can then be estimated using Equation (4). For a more comprehensive and detailed treatment of Gaussian process machine learning (GPML), the reader is referred to [34–38], and for the application of GPML application to spatiotemporally varying multiphase flows, the reader is referred to [17]. The methodology developed in this study can be applied to other areas of science and engineering with the notable change being the selection of separate kernel functions based on data behavior.

2.2. Framework

Figure 1 shows the schematic of the framework. It consists of three steps: (1) identification and generation of training and testing databases—these databases established using LES will be referred to as truth model hereafter; (2) training of the GPML algorithm using the training database; and (3) error quantification based on comparison of GPML predicted results with the truth model under test conditions. It should be noted that the testing database was not used for training, instead, it was used exclusively to test the accuracy of the GPML predictions at test points. Furthermore, the framework is agnostic to the source of data—the truth model can also be high-resolution experimental measurements. These three steps are described briefly below:

Step 1: generation of training (and testing) data

This step involves sampling the design space using a single variable worst-case scenario Latin Hypercube Sampling (LHS) technique [39,40] in combination with knowledge of physical processes dominating the physics under consideration. Since training datasets are expensive to generate (both computationally and experimentally), one of the goals was to use a minimum set of data points while maximizing accuracy, which was one of the reasons for choosing GPR which is known to do well with sparse data. Once a set of conditions for training data points was selected, LES-based calculations were conducted under each of those conditions. The configuration chosen here was representative of fuel injected into crossflowing air in a ramjet engine. Some of the relevant non-dimensional numbers for the present flow were the gas Reynolds number Re_g, Reynolds number based on jet diameter Re_d, $We_l = \rho_l U_R a/\sigma$ the liquid Weber number, $We_g = \rho_g U_R d/\sigma$ the gas Weber number, and momentum flux ratio q. The importance of some of these non-dimensional numbers for the present flow configuration is discussed in [32]. While we could have chosen to vary any of these important parameters to demonstrate that our model provides speed up and is accurate for turbulent multiphase flow problems, we chose the momentum flux ratio because it takes into account both the dynamics of the gaseous and liquid phase effects [16]. The penetration depth of the injected fuel depends upon the jet velocity. Momentum flux ratio was, therefore, chosen as a parameter of engineering significance, and the data set was collected at an increasing momentum flux ratio. As shown in our previous paper [17], adding more training points can increase the accuracy of predictions up to a certain extent, after which the accuracy improves at a much lower rate as compared to the decline in speedup with the addition of more points. To maintain a balance between speedup and accuracy, in this manuscript, we deemed a 15% error from truth data as appropriate to achieve two orders of magnitude accuracy over LES calculations. To test the accuracy of the trained algorithm, a few testing points were also determined and corresponding data were generated using LES.

Step 2: Gaussian process regression

This step involves selecting a kernel function to model the covariance of each of flow quantities of interest—velocities, turbulent kinetic energy (TKE), pressure, temperature, vapor fractions, liquid jet

penetration depth, and the Sauter mean diameter (SMD) of the droplet size distribution. In general, selecting kernel shapes is the most critical step in GP based prediction techniques. For the current work, kernel functions were selected based on the behaviors of the corresponding flow variables in the most dynamic regions of the flow domain, such as the wakes downstream of the liquid jet and the bluff body to ensure that maximum variation across the data points was captured. It should be noted that the "most" dynamic regions that we referred to in the previous sentence can be different for different variables. In addition to regions of the most action, since our goal was to have one kernel function to model the behaviors of a given flow variables at every point in the flow, the additive property of GP kernels facilitates their selection, such that different terms can dominate the model at different spatial locations. Another strategy can be to divide the domain into multiple zones based on some criteria (physics-based preferred), and then select different kernels for different zones. After selecting a kernel function, the entire flowfield is subjected to the GPR-based training process.

Step 3: error quantification

In this step, the accuracy of the machine learning algorithm was established using the L_1 error norm. This was conducted by predicting the gaseous and spray dynamics under testing conditions established in step 1 using the GPML algorithm trained in step 2 and comparing the results with the truth model. For clarity, the predictions from the trained machine learning algorithm will be referred to as predictions, and the results from LES calculations will be referred to as the truth model for the rest of this manuscript.
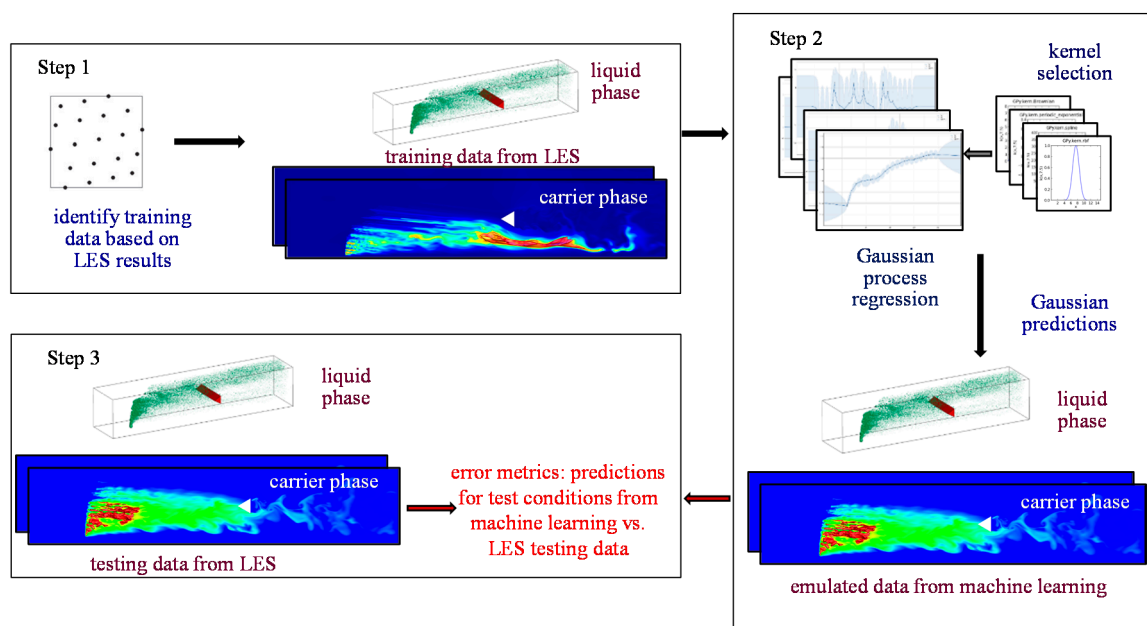
Figure 1. Outline of the Gaussian process-based machine learning (GPML) algorithm as applied to statistically stationary vaporizing turbulent liquid jet in crossflow datasets.

3. Theoretical Formulations and Governing Equations for the Truth Model

The truth model is based on a compressible Eulerian–Lagrangian LES framework. The next few subsections briefly describe the governing equations and the numerical methodology for the gaseous and the liquid phases. For a detailed description, the reader is referred to our previous publications [32,33].

3.1. Governing Equations for the Gaseous Phase

The governing equations were formulated and solved to satisfy the conservation of mass, momentum, energy, and species transport in Cartesian coordinates. Source terms appeared due to interaction with the dispersed phase droplets. Large eddy simulation (LES) technique was used to achieve turbulence closure. The Favre averaged equations are given by:

$$\frac{\partial \overline{\rho}}{\partial t} + \frac{\partial \overline{\rho}\widetilde{u}_i}{\partial x_i} = \widetilde{\dot{\rho}}_s \tag{6}$$

$$\frac{\partial \overline{\rho}\widetilde{u}_i}{\partial t} + \frac{\partial \overline{\rho}\widetilde{u}_i\widetilde{u}_j}{\partial x_i} = -\frac{\partial \overline{p}}{\partial x_i} + \frac{\partial (\overline{\tau}_{ij} - \tau_{ij}^{sgs})}{\partial x_i} + \widetilde{\dot{F}}_{s,i} \tag{7}$$

$$\frac{\partial \overline{\rho}\widetilde{E}}{\partial t} + \frac{\partial [(\overline{\rho}\widetilde{E} + \overline{p})\widetilde{u}_i]}{\partial x_i} = \frac{\partial (-\overline{q}_i + \widetilde{u}_j\overline{\tau}_{ji} - \sigma_i^{sgs} - H_i^{sgs})}{\partial x_i} + \widetilde{\dot{Q}}_s \tag{8}$$

$$\frac{\partial \overline{\rho}\widetilde{Y}_k}{\partial t} + \frac{\partial (\overline{\rho}\widetilde{Y}_k\widetilde{u}_i)}{\partial x_i} = \frac{\partial (-\overline{\rho}\widetilde{Y}_k\widetilde{V}_{i,k} - Y_{i,k}^{sgs} - \theta_{i,k}^{sgs})}{\partial x_i} + \widetilde{\dot{w}}_k + \widetilde{\dot{S}}_{s,k} \tag{9}$$

Spatially filtered variables, Favre-averaged variables, viscous stress tensor, and heat flux vector are denoted as: $\overline{f}$, $\widetilde{f} = (\rho f / \overline{\rho})$, τ_{ij}, $\overline{q}_i$, respectively. Y_k refers to the mass fraction of the kth species. The species diffusion flux is denoted by J. The sub-grid scale (SGS) stresses term τ_{ij}^{sgs}, SGS energy fluxes term H_i^{sgs} and SGS species fluxes $\theta_{i,k}^{sgs}$ result from filtering these convective terms. The SGS viscous work term, σ_{ij}^{sgs}, comes from correlations of the velocity field with the viscous stress tensor, and the SGS species diffusive fluxes term, $\theta_{i,k}^{sgs}$, comes from correlations of the velocity field with the species mass fractions with the diffusion velocities. The resolved-scale species mass production rate, $\overline{\dot{w}}_k$, is also unclosed. These SGS terms are closed using the algebraic version of the Smagorinsky model, as given by Erlebacher, et al. [6]. The details of the model are discussed elsewhere [41]. The terms $\widetilde{\dot{\rho}}_s$, $\widetilde{\dot{F}}_{s,i}$, $\widetilde{\dot{Q}}_s$, and $\widetilde{\dot{w}}_k$ represent the interphase exchange of mass, momentum, energy, and species between the gas phase and the spray. The coupling between the two phases is described in Section 3.2.2.

3.2. Dispersed Phase Modeling

A Lagrangian approach was used to model the droplets and their kinematics. The source terms associated with the transfer of mass, momentum, and energy between the dispersed and continuous phase were distributed across the diameter of the droplet. In this way, the present method accounted for the finite size of the droplets; details of this method can be found elsewhere [33]. Since tracking every droplet is a computationally expensive process, multiple droplets of similar size, location, velocity, and temperature were grouped together and treated as a parcel—this is known as the blob approach in the literature. The motion of particles was determined as Newton's second law of motion as follows:

$$\frac{dx_d}{dt} = u_d \tag{10}$$

$$m_d \frac{du_d}{dt} = F_d \tag{11}$$

where x_d is the instantaneous particle location, and $\dot{m}_d$ is the droplet mass. Only the force arising from skin friction and form drag was considered, after neglecting contributions from virtual mass, buoyancy, Basset forces, gravity, and lift.

$$F_d = \frac{1}{8}C_d\rho\pi d_d^2|u_R|u_R \tag{12}$$

d_p and u_R are the particle diameter and velocity relative to the surrounding carrier fluid, respectively. Droplet mass transfer due to vaporization is governed by the droplet continuity equation, as shown below:

$$\frac{dm_d}{dt} = -\dot{m}_d = \dot{m}^s \tag{13}$$

where $\dot{m}_d$ md is the mass of the droplet, which is given by $\frac{4}{3}\rho_d\pi r_d^3$, and $\dot{m}_d(>0)$ is the net mass transfer rate of droplet mass due to vaporization, which is expressed as:

$$\frac{\dot{m}_d}{\dot{m}_{d,\mathrm{Re}_{d=0}}} = 1 + \frac{0.278\mathrm{Re}_d^{1/2}Sc^{1/3}}{\left[1 + \frac{1.232}{\mathrm{Re}_d Sc^{4/3}}\right]^{1/2}} \tag{14}$$

$\mathrm{Re}_d = 0$ is the Reynolds number for a droplet at rest. The energy equation was solved to calculate the droplet heat transfer. This accounted for convective heat transfer between the droplet and surrounding air as well as latent heat of vaporization. Temperature was assumed to be uniform throughout the droplet. The equation governing the droplet temperature is given by:

$$m_d C_l \frac{dT_d}{dt} = \dot{Q}_{conv} - \dot{m}_d L_v = h_d \pi d_d^2 (\widetilde{T} - T_p) - \dot{m}_d L_v \tag{15}$$

where $\dot{Q}_{conv}$ is the convective thermal energy transfer rate, C_l is liquid heat capacity, h_d is the heat transfer coefficient, and L_v is the latent heat of vaporization.

3.2.1. Spray Closure Modeling

The hydrodynamic stability analysis developed by Reitz [42] was implemented to model the jet injection and atomization process. Here, the liquid jet was injected as blobs of diameter equal to injector exit diameter d_o, i.e., $d_{int} = d_o$. The number of blobs injected per unit time was based on the injector mass flow rate. Each blob was injected at velocity U_o and at a spray angle θ, which was assumed to be uniformly distributed between 0 and Φ, with:

$$V_0 = U_0 \tan(\theta/2) \tag{16}$$

$$\tan(\theta/2) = (A_1 \Lambda \omega / U_0) \tag{17}$$

where A_1 has a value of 0.188 for a sharp entrance nozzle with a length to diameter ratio of 5. The theory of Kelvin–Helmholtz (K-H) instability in jets was used to determine the frequency associated with the fastest growing instability wave, Λ, and the corresponding wavelength, λ. These values were determined by performing a curve-fit solution of the linearized hydrodynamic equations, and presented in non-dimensional quantities which are specified by:

$$\frac{\Lambda}{a} = 9.02 \frac{\left(1 + 0.45Z^{0.5}\right)\left(1 + 0.4T^{0.7}\right)}{\left(1 + 0.87We_g^{1.67}\right)^{0.6}} \tag{18}$$

$$\Omega\left(\frac{\rho_l a^3}{\sigma}\right)^{0.5} = \frac{\left(0.34 + 0.38We_g^{1.5}\right)}{(1 + Z)(1 + 1.4T^{0.6})} \tag{19}$$

where $Z = We_l^{0.5}/\mathrm{Re}_l$ is the Ohnesorge number, $We_l = \rho_l U_R a/\sigma$ is the liquid Weber number, $We_g = \rho_g U_R d/\sigma$ is the gas Weber number, and $T = ZWe^{0.5}$ is the Taylor number. The variable a is the blob/parent droplet radius. The subscripts l and g denote the liquid and gas phases, respectively. The breakup of liquid droplets moving in surrounding air was assumed to be caused by instability waves induced by the deceleration of the droplets caused by aerodynamic drag. The breakup of the

droplets was modeled such that their sizes were proportional to the wavelength of the fastest growing instability wave droplets, and is given by:

$$r = B_0\Lambda \quad (B_0\Lambda \leq a) \tag{20}$$

$$r = \min\left[\left(3\pi a^2 W/2\Omega\right)^{0.33}, \left(3a^2\Lambda/4\right)^{0.34}\right] \quad (B_0\Lambda > a, \text{one time only}) \tag{21}$$

where B_0 is constant and equals 0.61. Once the parent droplet broke up, its mass was assigned to child droplets, and then these droplets were tracked. The rate of change of drop radius in a parent parcel was assumed to obey the following equation:

$$\frac{da}{dt} = -\frac{(a-r)}{\tau} \tag{22}$$

where $\tau = 3.726B_1 a/\lambda\omega$, with B_1 being a breakup time constant, and is set to a value of 1.73 as per the recommendation made by Liu et al. [43]. Secondary atomization was modeled by using The Taylor analogy breakup (TAB) model [44], and the impingement of droplets on solid surfaces (walls) was treated using the model developed by Mundo et al. [45]. More information about multiphase flow modeling can be found in [33].

3.2.2. Coupling between Dispersed and Carrier Phase

Coupling between the dispersed phase and gas phase was accounted for by calculating the interphase exchange between the dispersed phase and carrier air. The source terms that accounted for this exchange were computed as follows:

$$\begin{pmatrix} \widetilde{\dot{\rho}}_s \\ \widetilde{\dot{F}}_s \\ \widetilde{\dot{Q}}_s \\ \widetilde{\dot{S}}_{i,s} \end{pmatrix} = \begin{pmatrix} \sum\limits_m n_d \dot{m}_d \\ \sum\limits_m n_d\left[\dot{m}_d u_d - \frac{4\pi}{3}\rho_d r_d^3 \frac{du_d}{dt}\right] \\ \sum\limits_m n_d\left[\dot{m}_d h_{v,s} - h_d \pi d_d^2(T - T_d) - u_d m_d \times \frac{du_d}{dt} + \dot{m}_d\left(\frac{1}{2}u_d \times u_d\right)\right] \\ \sum\limits_m n_d \dot{m}_d \end{pmatrix} \tag{23}$$

where the summation index m accounts for all the droplet parcels that cross a given computational cell volume.

Governing equations for the gaseous phase was solved using the finite volume method employed on a structured grid. The spatial discretization method used was a second-order accurate central-difference scheme used in generalized coordinates. A fourth-order matrix dissipation scheme with a total variation-diminishing method was applied to ensure numerical stability by preventing numerical oscillations in locations where steep gradients may be present. Temporal discretization was performed using the fourth order Runge–Kutta method. Multi-block domain decomposition was used to implement parallel computations, with message passing interface at the domain interface. Further details of these techniques can be found elsewhere [1,33,41].

4. Error Quantification and Speedup

Errors were estimated to quantify the degree of accuracy achieved with the surrogate model when compared with the original simulation. The error between the predicted quantity and truth model was estimated for each grid location and averaged for all grid points to obtain the average L_1 error defined by:

$$\text{error} = \frac{1}{n}\sum_{i}^{n}\left(\frac{|\text{truth} - \text{emulation}|}{\text{truth}}\right)_i \tag{24}$$

The speedup afforded by the surrogate model was measured as the ratio of CPU hours needed by the truth model and the GPML algorithm for the test condition under consideration. The number of processors required and the actual computational time in hours were combined to arrive at a CPU hours metric as shown below:

$$\text{CPU hours} = (\text{Number of processors} \times \text{Physical time in hours}) \tag{25}$$

The speedup factor could then be estimated by:

$$\text{speedup factor} = \left(\frac{\text{CPU hours}_{\text{truth}}}{\text{CPU hours}_{\text{predicted}}} \right) \tag{26}$$

5. Results and Discussion

The results are organized into four subsections. In the next subsection, comparisons between the LES results and experimental measurements are presented to establish the credibility of the truth model. A brief description of flow features associated with vaporizing liquid jet in crossflow is also provided. This is followed by a subsection describing the selection of training points. Results for the gaseous and liquid phases predicted by GPML are discussed in Sections 5.3 and 5.4, respectively.

5.1. Experimental Validation of the Truth Model

The theoretical framework presented above is validated with experimental measurements made by Stenzler et al. [21]. The results were compared for both vaporizing and non-vaporizing cases. Figure 2 shows the schematic of the computational setup. Table 1 lists all the operating conditions, the physical properties of air as well as the liquid jet, and all relevant non-dimensional numbers used here.
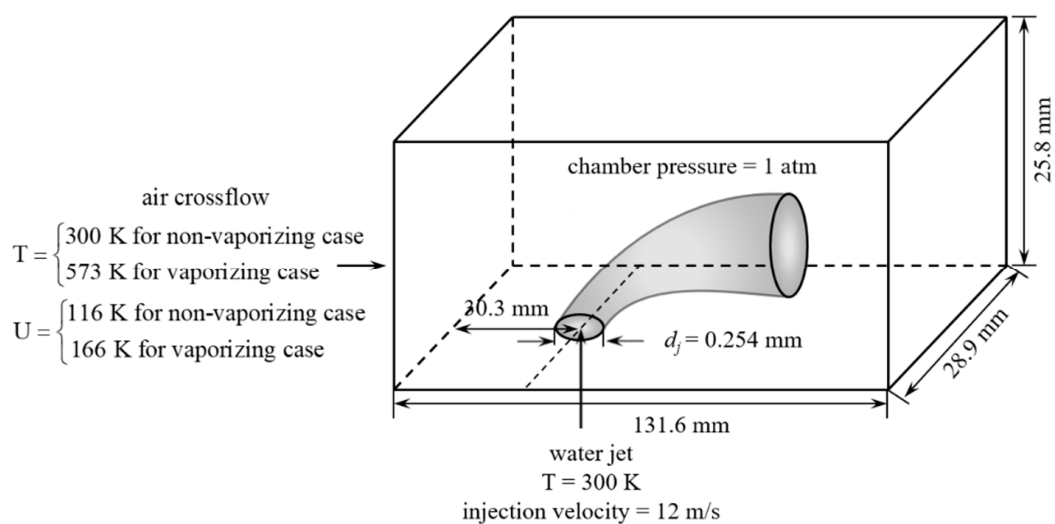
Figure 2. Schematic of the computational setup for the large eddy simulation (LES) model validation. Two cases were conducted for transverse water injection in air crossflow. For the non-vaporizing case, the air temperature was 300 K and for the vaporizing case, it was 573 K.

Table 1. Operating conditions, physical properties, and relevant non-dimensional numbers used for model validation cases.

	Non-Vaporizing Case	Vaporization Case
T_a (K)	300	600
U_a (m/s)	116	166
U_j (m/s)	12.01	12.3
ρ_a (kg/m^3)	1.18	0.62
σ_o (N/m)	7.28×10^{-2}	7.28×10^{-2}
μ_a (Ns/m^2)	1.86×10^{-5}	2.98×10^{-5}
q	9	9
We	68	68
Re	2059.3	933.7

A total of 5.1×10^6 grid points were used to mesh the entire computational domain. Figure 3 shows a perspective view of the multi-block grid that was used. However, only every 5th grid volume is retained in the figure. The grid was particularly well refined close to the jet injection location so that the unsteady wake shed downstream may be well resolved. The smallest grid size was 0.02 and 0.01 mm in the x and z directions, respectively. At least 10 grid points were used to resolve the diameter of the jet. The dense grid was retained up to 12 injector diameters (3 mm) downstream of the injector location, and 1.6 diameters (0.4 mm) in the z direction. The grid was then gradually stretched before refining it once again, close to wall surfaces. The current grid resolved over 80% of the turbulent kinetic energy (TKE) for a jet Reynolds number of 2059. To account for inflow turbulence reported in experiments, a precursor turbulent channel flow simulation was first conducted at the appropriate Reynolds number to generate the corresponding turbulent flowfield. The instantaneous flowfield data obtained from this turbulent channel flow were then fed into the computational domain for the validation cases.
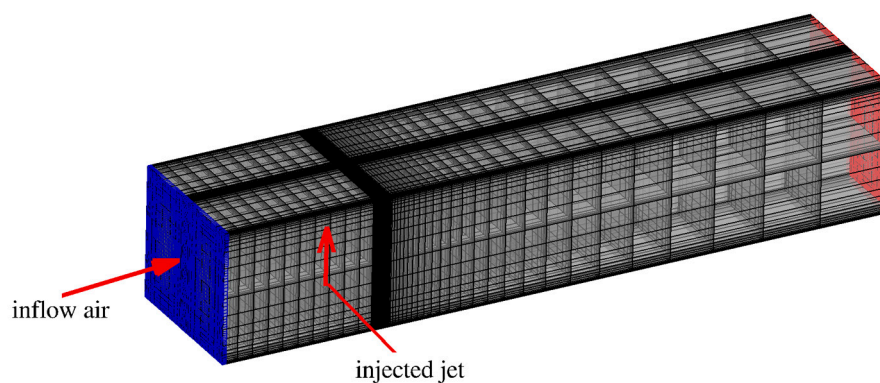

inflow air

injected jet

Figure 3. Grid for the validation cases. A total of 5.1×10^6 grid points were used to mesh the computational domain. This figure shows every 5th grid volume.

Figure 4a,b show the comparison of the liquid jet penetration and its trajectory obtained from the present study. Results were compared with correlations developed by Stenzler et al. [21], based on experimental measurements. The correlation curves represent the upper edge of the jet core. The outer edge of the jet is based on experimental imaging and is the locus of points where the droplet/jet was observed at the furthest point in the direction perpendicular to the air flow. The results are shown for the non-vaporizing and vaporizing cases, respectively, showing good agreement with the correlation curves.
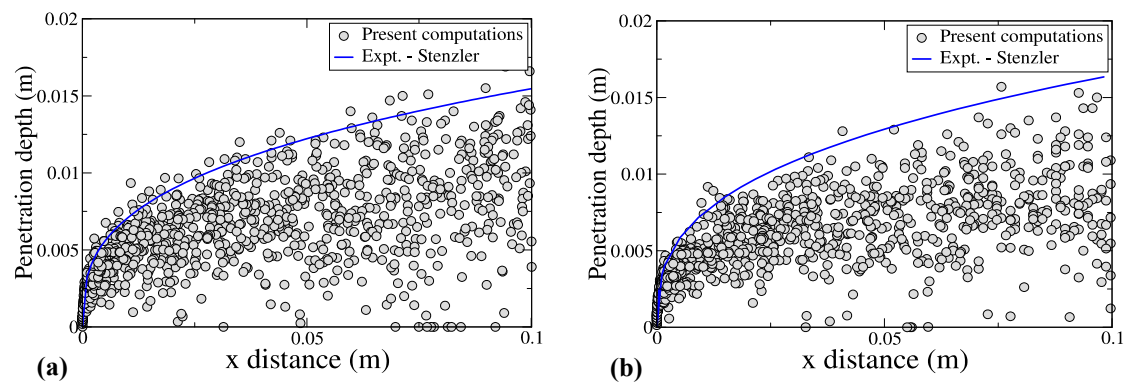
Figure 4. Comparison of liquid jet penetration with experimental measurements for (**a**) non-vaporizing and (**b**) vaporizing cases.

A quantitative comparison was made by showing droplet size distributions in Figure 5a–d, where the evolution of the Sauter mean diameter (SMD) with distance is shown by presenting the SMD contours at locations 1,9,18, and 25.4 mm downstream of the jet. The results shown here are for the non-vaporizing case. The gradual expansion of the jet core is evident from the figures. Upon further inspection, it is seen that the largest droplets (55–65 µm) were found close to the upper edge of the jet core, while the midsized droplets were concentrated in the jet core. Relatively smaller droplets were distributed close to the bottom. As shown in Figure 6, the SMD distribution compares well with experimental measurements made by Stenzler et al. [21], thus confirming our observations. The comparison was made for the location that was 25.4 mm downstream of the injector. While the trends were generally in agreement, the experimental measurements show that there was a small fraction of larger sized droplets at the center of the core as compared to the present simulations that predict these droplets to be presented mostly close to the upper edge of the jet core.
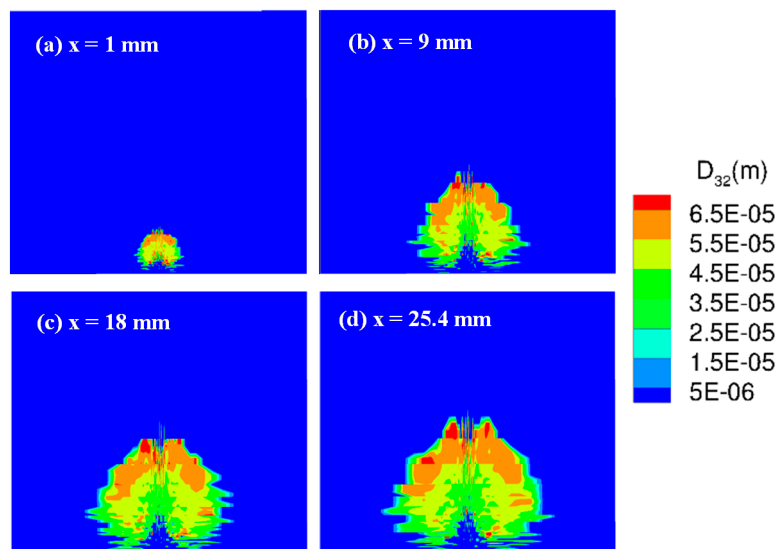
Figure 5. Sauter mean diameter (SMD) distribution on the y–z plane: progressive streamwise locations showing the development of the spray field at Ta = 300 K at (**a**) 1 mm, (**b**) 9 mm, (**c**) 18 mm, and (**d**) 25.4 mm downstream of the injection location.

(a) present computation   (b) experiment
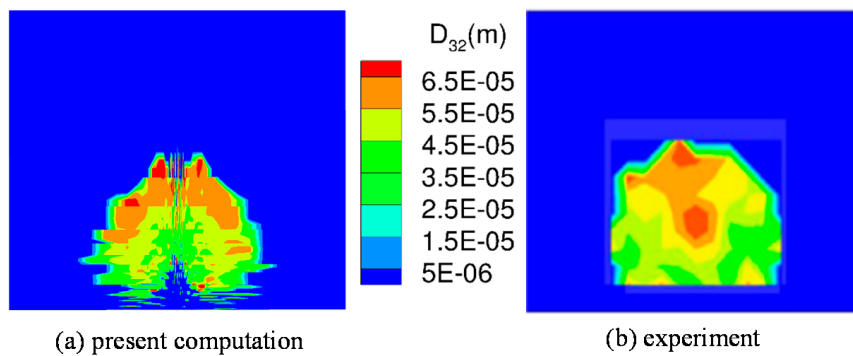
Figure 6. SMD distribution 25.4 mm downstream of the injector; (**a**) present calculations and (**b**) experiments of Stenzler et al. [21].

It should be noted that the wake behind the liquid jet needs to be captured with a denser grid to ensure the accuracy of both the gaseous and liquid spray fields and for reliable droplet statistics. The unsteady wake was resolved using the finite-sized formulation of the two-way coupling developed by Khare et al. [33]. Figure 7 shows the instantaneous contour of vorticity magnitude immediately downstream of the jet wake. An intense exchange of momentum between the liquid jet and inflowing air gave rise to turbulence unsteady fluctuations in the wake of the jet. These turbulent fluctuations constituted shed wake vortices, counter-rotating vortex pairs, and vortices generated in the boundary layer close to the bottom wall. Furthermore, to quantitatively identify the turbulent wake, time-series data for velocity fluctuations were collected at a probe located 2 mm immediately downstream of the liquid injection location. Figure 8 shows the spectra obtained by taking the fast Fourier transform (FFT) of the velocity fluctuations, which follows the $-5/3$ law of the Kolmogorov–Obukhov theory for turbulent flows.
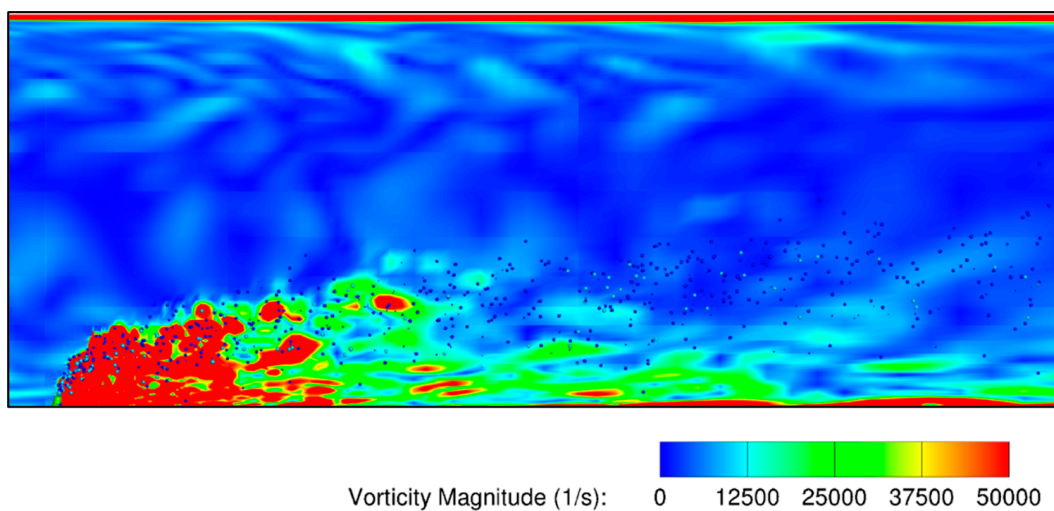

Figure 7. Instantaneous vorticity contour showing the turbulent wake formation downstream of the liquid jet.
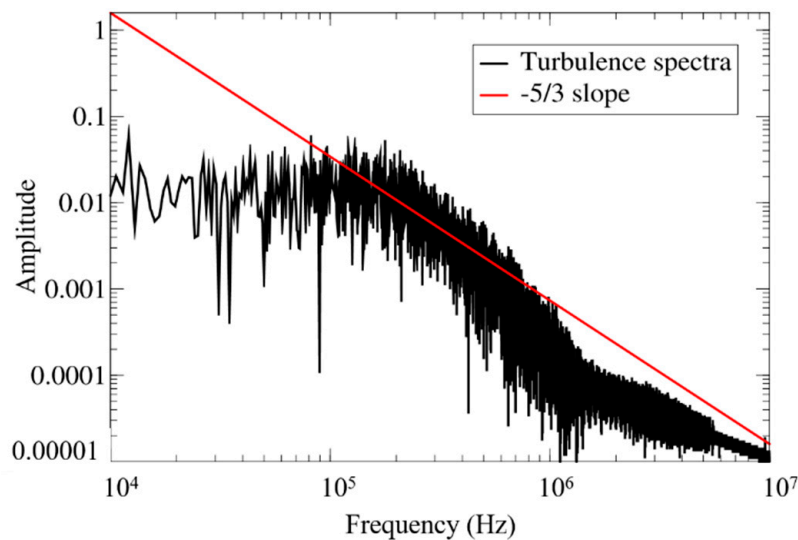
**Figure 8.** Turbulence spectra 2 mm downstream of the injector showing the −5/3 slope.

5.2. Configuration and Operating Conditions for Training Dataset

A more realistic configuration relevant to ramjet and afterburner engines was used to demonstrate that our machine learning algorithm can be employed for design space exploration parametrized by the momentum flux ratio that ranges from 5 to 120; all other relevant non-dimensional numbers including the Reynolds and Weber numbers are approximately constant. Figure 9 shows a schematic of the configuration—the major difference from the validation case was the presence of a triangular bluff body downstream of the liquid injection location. The flow configuration consisted of liquid water injected into cross-flowing air at atmospheric pressure but at an elevated temperature of 600 K. The computational domain consisted of a $248.7 \times 50 \times 50$ mm^3 domain. The liquid jet was located at L_j = 40 mm from the leading edge of the domain, and the v-gutter was located further downstream at a separation of L_s = 100 mm. The v-gutter is an equilateral triangle with a side of length of L = 10 mm. Airflow was kept constant at a flow velocity U_g of 141.6 m/s, and the momentum flux ratio was varied between 5 and 120 by increasing the liquid jet injection velocity U_l from 7.04 to 37.31 m/s. It should be noted that this led to an increase in the Weber number by 6%. Quantitative comparisons between the GPML predictions and truth models were made at four planes corresponding to x = 80, 120, 170, and 210 mm from the inlet. Knowledge of the behaviors of flow quantities such as velocities, pressure, temperature, vapor-fraction, and turbulent kinetic energy (TKE) with increasing momentum flux ratio was utilized for identifying a set of training and testing points. Amounts of 420, 198, and 101 grid points were used in the x, y, and z-directions, respectively, with 240 grid points upstream of the bluff body and 180 downstream of it. This results in a total of 8 million points to mesh the computational grid.
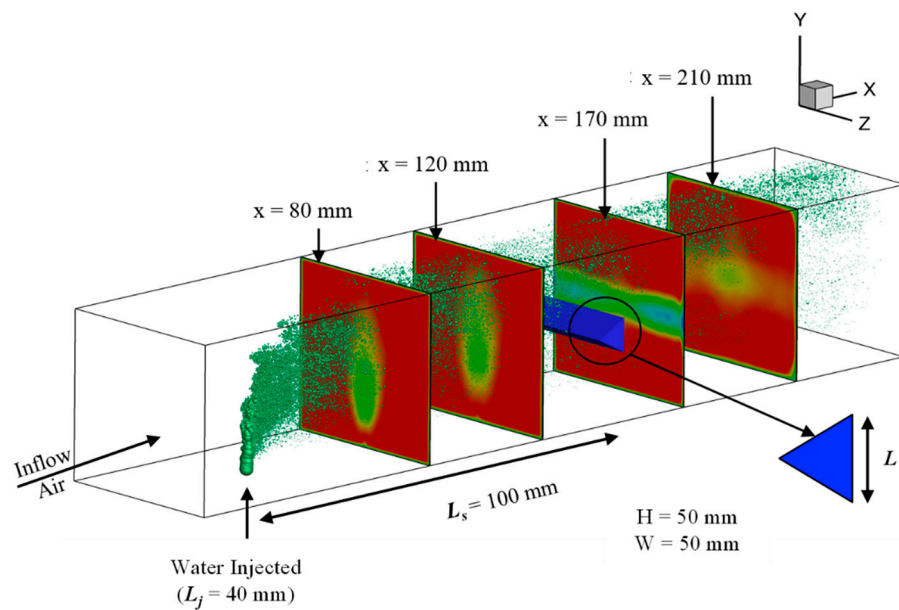
Figure 9. Schematic of the configuration used to demonstrate design space exploration of vaporizing liquid jet in crossflow using GPML. Quantitative comparisons between the GPML predictions and truth models were made at four planes corresponding to x = 80, 120, 170, and 210 mm from the inlet.

Figure 10 shows the spray fields obtained from LES-based calculations at three representative momentum flux ratios of q = 5, 30, and 120. At the value of q = 5, the jet did not have sufficient momentum to penetrate deep into the test section, and the droplets were concentrated closer to the bottom wall, well below the v-gutter. At q = 30, the core of the spray field penetrated deeper and was within the vicinity of the v-gutter. With a further increase in momentum flux ratio, for q = 120, the injected jet impinged on the upper wall and most of the droplets were concentrated close to the upper wall.
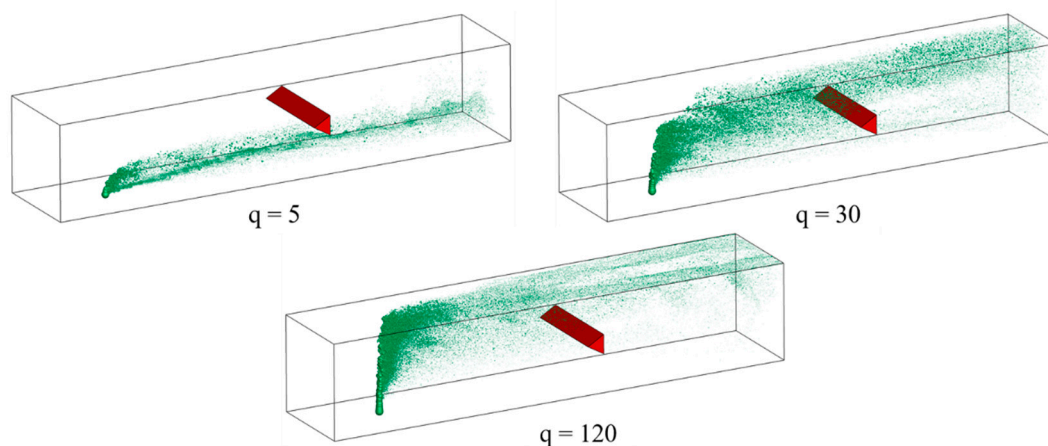
Figure 10. Instantaneous spray field for three representative cases of momentum flux ratios of 5, 30, 120 showing qualitative differences in the spray field.

The design of experiments based on LHS worst-case scenario and the domain knowledge of jet behavior facilitated the selection of training data points. Both the extreme cases of q = 5, 120, and an intermediate point q = 10 where the jet impinged on the v-gutter, were included in the training dataset of the GPML algorithm. Table 2 lists the points selected for training the surrogate model. Once the

algorithm was trained, several operating points in the momentum flux parameter space were chosen at random to test its accuracy; these testing points are listed in Table 3.

Table 2. List of momentum flux ratio values for training the machine learning emulator.

Training Dataset Identified by the Momentum Flux Ratio, q						
5	8	10	30	50	100	120

Table 3. List of momentum flux ratio values to test the trained GPML algorithm.

Testing Dataset Identified by the Momentum Flux Ratio, q						
7	9	20	40	65	90	110

As mentioned before, to quantify the accuracy, we compared the GPML predictions at the mid-plane of the computational domain for time-average streamwise velocity, TKE, pressure, temperature, the mass fraction of vaporized water, and temperature are reported at four select streamwise stations located at 80, 120, 170, and 210 mm from the inlet plane under test conditions listed shown in Figure 9. The first station was in the wake of the liquid jet, and the second one further downstream before the v-gutter. The next two locations were chosen downstream of the v-gutter, with one within the wake of the v-gutter and the other one further downstream of the wake.

Once trained, the machine algorithm was capable of emulating any momentum flux ratio within the training bounds of q = 5 and 120. In the current manuscript, we show the comparison of the predictions by the trained algorithm with truth data for 2 testing points identified by q = 7 and 40.

5.3. Results of Machine Emulation

In the next few subsections, we will discuss how the statistically stationary time-averaged flow variables predicted by the trained GPML algorithm compared with the truth data obtained using LES at test points of q = 7 and 40 at the mid-plane followed by the profiles of flow variable at the previously mentioned streamwise locations. The errors will be quantified by L_1 norms. While a brief description of the physical processes is provided as we discuss each flow variable, readers are referred to our previous work that discusses the effect of the momentum flux ratio on the detailed flow physics [32]. It should be noted that the accuracy of the GPML algorithm is directly proportional to the accuracy of the training data, which in this case is the results from LES calculations.

5.3.1. Velocity Magnitude

Figure 11a shows the mid-plane contours of velocity magnitude predicted by the GPML algorithm (left) and truth data (right) for test conditions of q = 7 and 40, showing excellent qualitative agreement. To further quantify differences, Figure 11b shows a comparison of the velocity profiles at x = 80, 120, 170, and 210 mm—the solid lines represent the truth model and the dashed lines show machine learning emulations. The wakes downstream of the liquid jet and the v-gutter and the flow acceleration on the top and bottom of the v-gutter were accurately captured by the GPML predictions, as shown in Figure 11b at all four locations. To quantify the differences, the L_1 norm was calculated for the entire flow domain. The machine learning predictions for q = 7 differed from the truth model by 6.05% and for q = 40 by 6.76%. The kernel function used for machine emulation of velocity magnitude is given by:

$$\text{kernel} = (\text{Rational Quadratic} + \text{Spline} + \text{Brownian})$$

$$k(x, x') = \left[\left(1 + \frac{(x - x')^2}{2\alpha l^2} \right)^{-\alpha} + \left(\xi + xx' + \Gamma\left(\frac{1}{2}|x - x'|\min(x, x') + \frac{1}{3}\min(x, x')^3 \right) \right) + (\min(x, x')) \right] \quad (27)$$
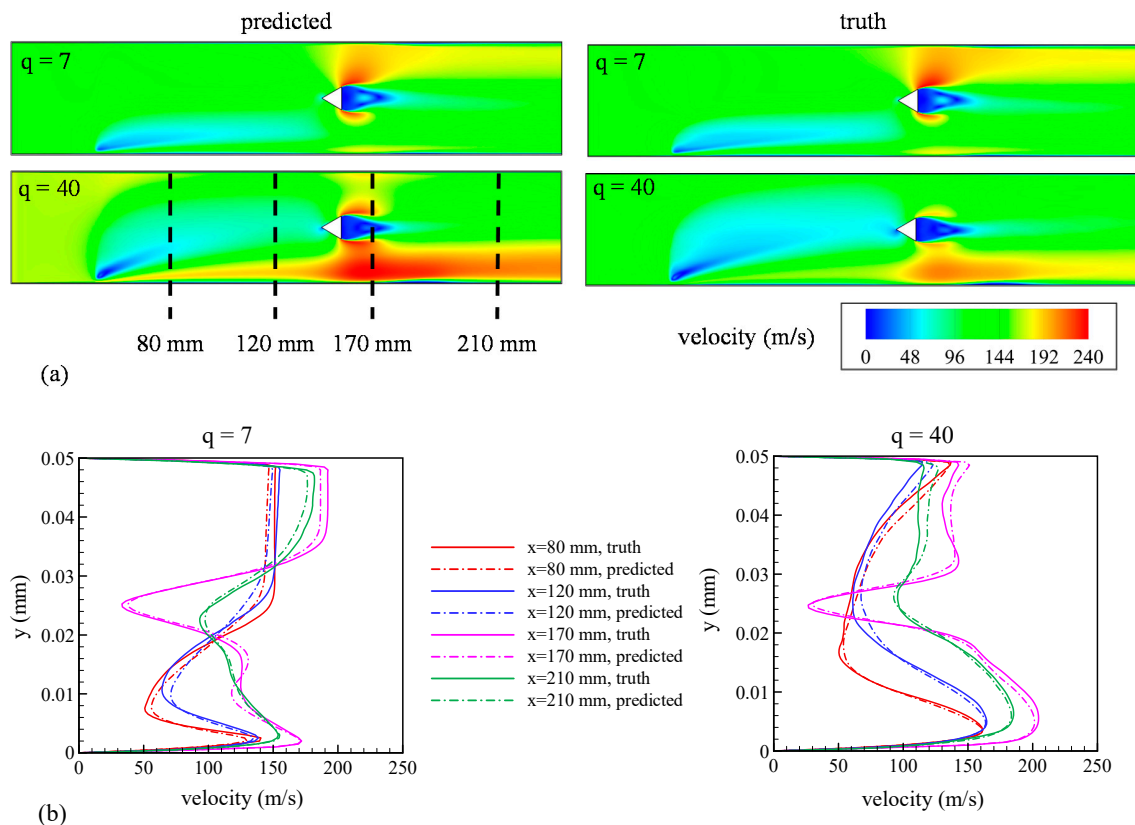
Figure 11. Comparison between predictions from the GPML algorithm (referred to as predictions) with LES results (referred to as truth) for test points of q = 7 and 40. (**a**) Contours of velocity magnitude in the mid-plane of the flow domain; (**b**) variation of velocity magnitude along the y-axis at x = 80, 120, 170, and 210 mm.
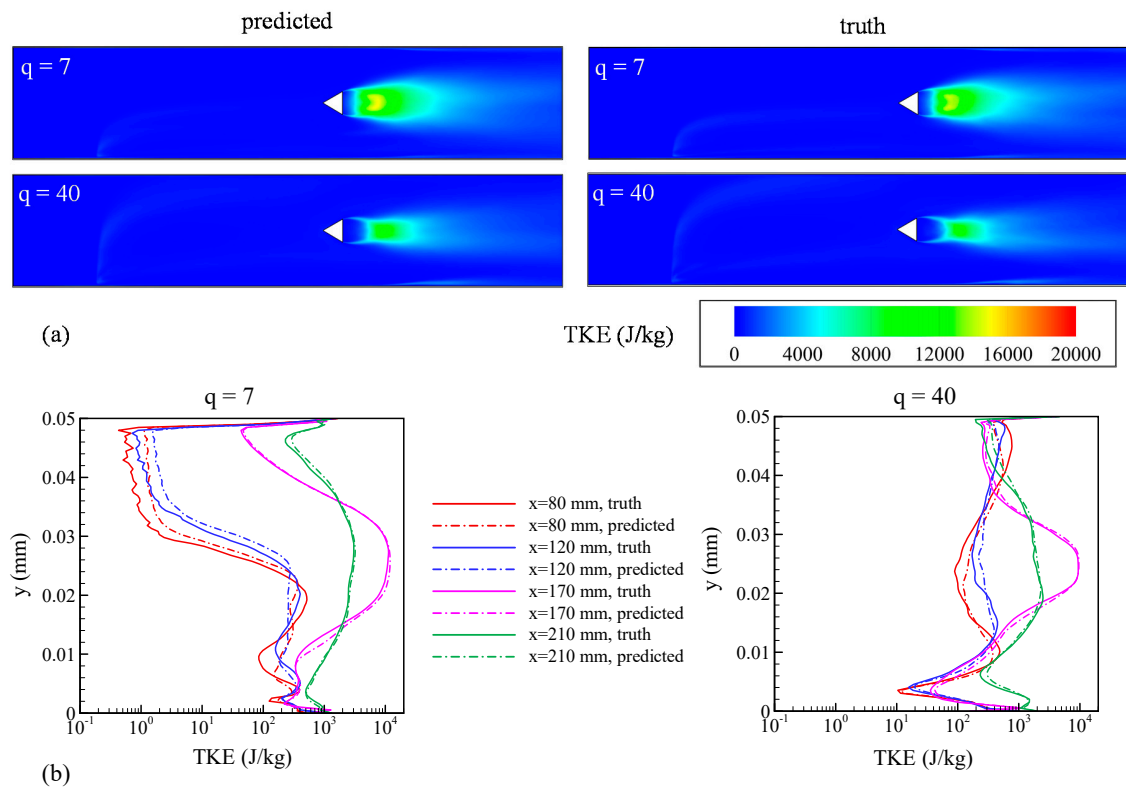
5.3.2. Turbulent Kinetic Energy

To demonstrate that our algorithm can predict second-order turbulence statistics, TKE was also predicted and compared with the truth data at test points, shown in Figure 12a,b. TKE contours provided an estimate of the level of turbulence at different locations in the flow domain. Momentum transfer from the gas phase to the dispersed liquid phase had a significant effect on TKE. Velocity fluctuations increased in the immediate wake of the jet but were still a few orders of magnitude lower than that of turbulent fluctuations in the unsteady wake of the v-gutter. As the flow accelerated past the bluff-body, these fluctuations were intensified resulting in higher TKE downstream of the v-gutter—turbulent fluctuations were the strongest at this location. This phenomenon can be observed in the TKE plots in Figure 12b, at x = 170 mm for both cases. For q = 40, the intensity of the turbulent fluctuations was reduced due to a higher transfer of momentum from the gas phase to the liquid phase. The unsteady wake was smaller in size when compared to the q = 7 case as the injected liquid phase damped some of these fluctuations, which resulted in reduced freestream velocity and turbulence intensity. All these observations were effectively captured by the GPML predictions not only qualitatively but also quantitatively presented in Figure 12b, where the solid lines represent the truth model and the dashed lines show machine learning emulations. The error quantified by the L_1 norm for the entire flow domain was 10.47% for q = 7 and 13.26% for q = 40. The kernel function used for machine emulation of TKE is given by:

$$kernel = (Exponential + Brownian + Spline \times (1 + Brownian))$$

$$k(x,x') = \left[\begin{array}{l} (\exp(-\theta|x-x'|)) + \min(x,x') \\ \qquad + \left(\xi + xx' + \Gamma\left(\frac{1}{2}|x-x'|\min(x,x') + \frac{1}{3}\min(x,x')^3\right)\right) * (1+\min(x,x')) \end{array} \right] \quad (28)$$

predicted

q = 7

q = 40

(a)

truth

q = 7

q = 40

TKE (J/kg)

0   4000   8000   12000   16000   20000

q = 7

q = 40

x=80 mm, truth
x=80 mm, predicted
x=120 mm, truth
x=120 mm, predicted
x=170 mm, truth
x=170 mm, predicted
x=210 mm, truth
x=210 mm, predicted

y (mm)

TKE (J/kg)

y (mm)

TKE (J/kg)

(b)

Figure 12. Comparison between predictions from the GPML algorithm (referred to as predictions) with LES results (referred to as truth) for test points of q = 7 and 40. (**a**) Contours of TKE in the mid-plane of the flow domain; (**b**) variation of TKE along the y-axis at x = 80, 120, 170, and 210 mm.
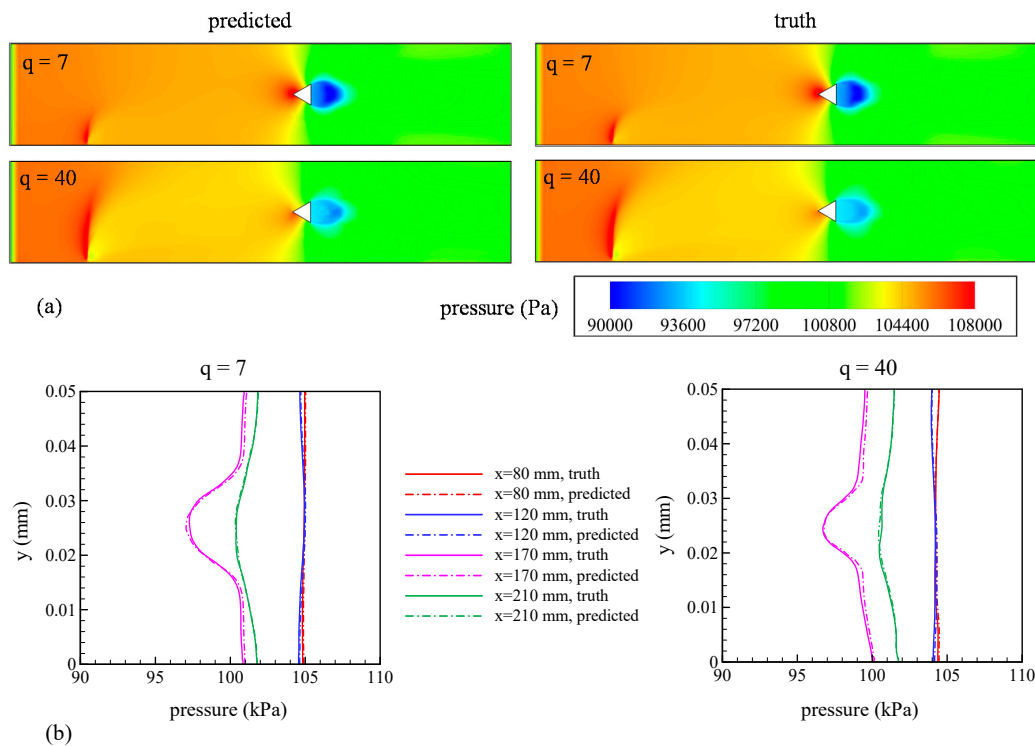
5.3.3. Pressure and Temperature

Figure 13a,b show a comparison of pressure fields predicted by the GPML algorithm and truth data from LES showing excellent agreement for q = 7 and 40, evidenced by the L_1 error of 0.05% for q = 7 and 0.07% for q = 40.

Figure 14a shows the temperature contours for both cases. For q = 7, since most of the injected liquid was in the lower half of the domain, a significant drop in temperature was observed in the jet wake—energy from the gas was used to vaporize the liquid jet. Similar trends were observed for q = 40, except that the lower temperature zone encompassed a considerably larger region of the domain in the y-direction because of the higher momentum carried by the liquid. As shown in Figure 14b, all these trends were effectively captured quantitatively by the GPML algorithm. The L_1 error for temperature over the entire domain was found to be 1.18% for both q = 7 and q = 40. The kernel function used for machine emulation of pressures and temperatures is given by:

$$\text{kernel} = (\text{RationalQuadriatic} + \text{Spline} + \text{Brownian})$$

$$k(x,x') = \left[\left(1 + \frac{(x-x')^2}{2\alpha l^2} \right)^{-\alpha} + \left(\xi + xx' + \Gamma\left(\frac{1}{2}|x-x'|\min(x,x') + \frac{1}{3}\min(x,x')^3\right) \right) + (\min(x,x')) \right] \quad (29)$$
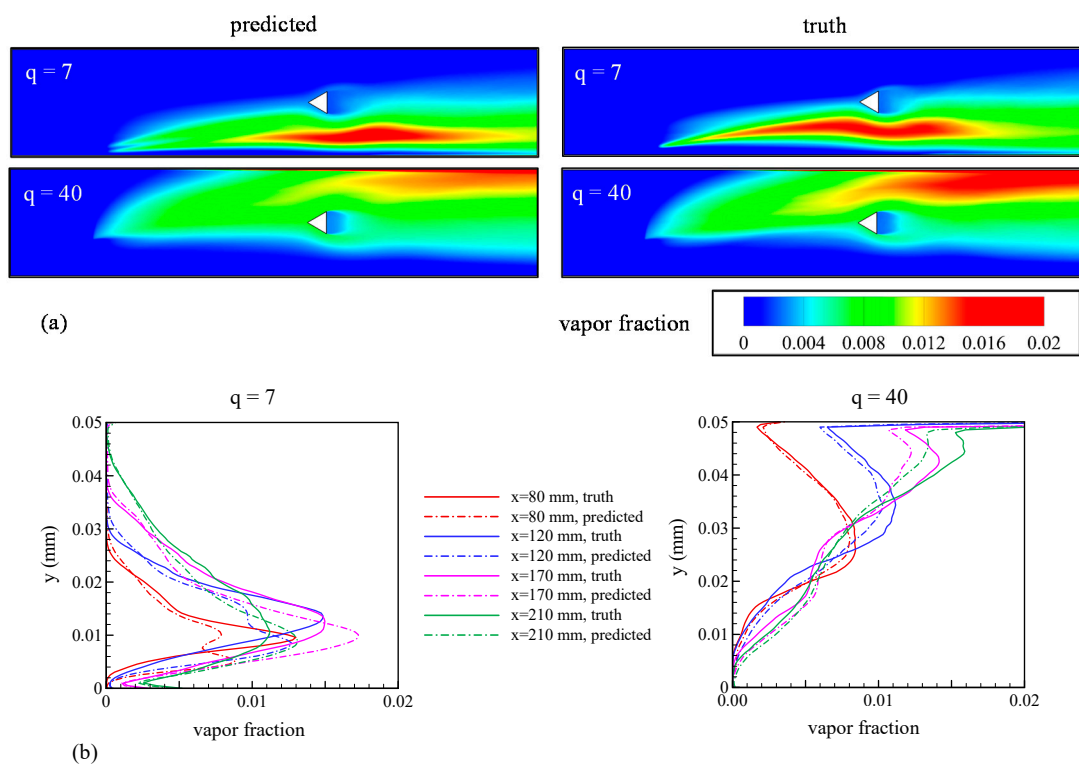
Figure 13. Comparison between predictions from the GPML algorithm (referred to as predictions) with LES results (referred to as truth) for test points of q = 7 and 40. (**a**) Contours of pressure in the mid-plane of the flow domain; (**b**) variation of pressure along the y-axis at x = 80, 120, 170, and 210 mm.
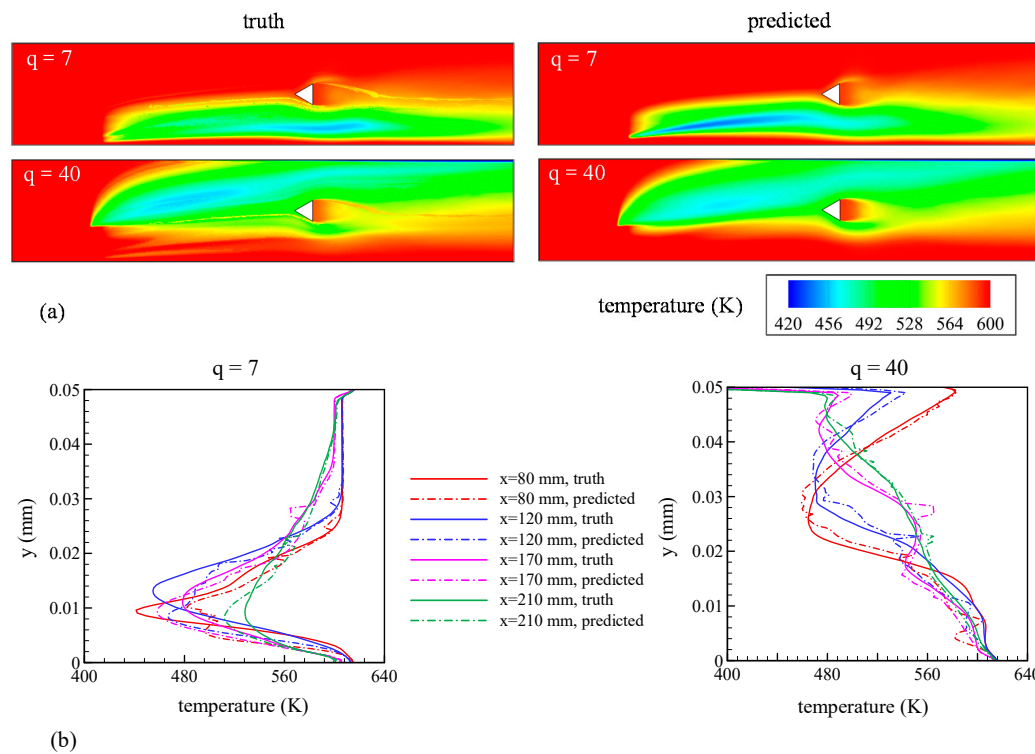
Figure 14. Comparison between predictions from the GPML algorithm (referred to as predictions) with LES results (referred to as truth) for test points of q = 7 and 40. (**a**) Contours of temperature in the mid-plane of the flow domain; (**b**) variation of temperature along the y-axis at x = 80, 120, 170, and 210 mm.

5.3.4. Liquid-Vapor Fraction

Heat transfer from the hot gas to the cooler water resulted in the vaporization of the liquid phase. Therefore, the variation of the liquid-vapor fraction closely corresponded to the locations of lower temperature regions discussed in the previous subsection. These are shown qualitatively as vapor fraction contours in Figure 15a and quantitatively at the four stations in the streamwise direction in Figure 15b. The L_1 error was calculated to quantify the accuracy of the GPML predictions as compared to the truth model; L_1 errors were found to be 11.91% for q = 7 and 7.99% for q = 40. The kernel function used for machine emulation of the liquid-vapor fraction is given by:

$$\text{kernel} = (\text{Exponential} + \text{Spline} + \text{Brownian})$$

$$k(x, x') = \left[(\exp(-\theta|x - x'|)) + \left(\xi + xx' + \Gamma\left(\frac{1}{2}|x - x'|\min(x, x') + \frac{1}{3}\min(x, x')^3 \right) \right) + (\min(x, x')) \right] \quad (30)$$
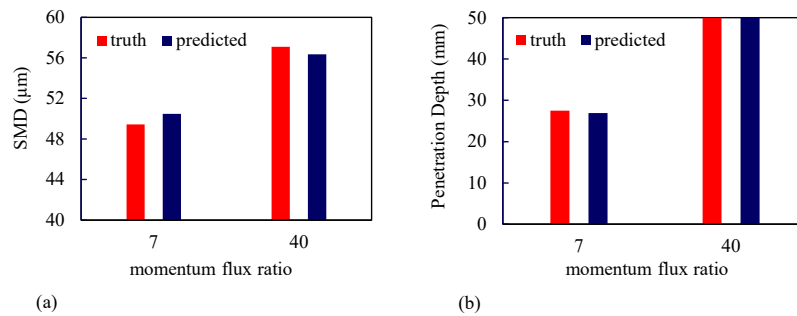

Figure 15. Comparison between predictions from the GPML algorithm (referred to as predictions) with LES results (referred to as truth) for test points of q = 7 and 40. (**a**) Contours of liquid vapor fraction in the mid-plane of the flow domain; (**b**) variation of liquid vapor fraction along the y-axis at x = 80, 120, 170, and 210 mm.

5.3.5. Sauter Mean Diameter (SMD) and Jet Penetration Depth

Figure 16a,b, respectively show a comparison of the Sauter mean diameter and liquid jet penetration depth predicted by the GPML algorithm and the truth model for test points corresponding to q = 7, and 40. The L_1 error for SMD was 2.11% for q = 7 and 1.30% for q = 40, and for the liquid jet penetration depth was 2.03% for q = 7 and 0.55% for q = 40. The kernel function used for machine emulation of SMD and jet penetration depth is given by:

$$\text{kernel} = (\text{Spline} + \text{Polynomial} + \text{Brownian})$$

$$k(x, x') = \left[\left(\xi + xx' + \Gamma\left(\frac{1}{2}|x - x'| \min(x, x') + \frac{1}{3}\min(x, x')^3 \right) \right) + \left(\alpha^2 x^{\mathrm{T}} x + \beta^2 \right) + \left(\min(x, x') \right) \right] \quad (31)$$

(a) (b)

Figure 16. Comparison between predictions from the GPML algorithm (referred to as predictions) with LES results (referred to as truth) for test points of q = 7 and 40. (**a**) SMD; (**b**) liquid jet penetration depth.

5.4. Speedup

The speedup was calculated as follows:

$$\text{speedup factor} = \left(\frac{\text{CPU hours}_{\text{LES}}}{\text{CPU hours}_{\text{GPML}}} \right) \quad (32)$$

The total CPU hours required by the LES-based calculations to reach a statistically stationary state and for collecting reliable higher-order statistics for the test points were approximately 27,360 h. A total of 21 h on 20 cores were needed to train the GPML algorithm, resulting in a total of 420 CPU hours. Thus, the speedup was approximately 65.

6. Summary

A data-driven surrogate model was successfully built by extending a previously developed Gaussian process-based algorithm to accommodate turbulent multiphase flows. The surrogate model was applied to liquid jet injection in turbulent air crossflow, a canonical configuration that represents fuel injection in afterburner and ramjet engines. The algorithm was trained using data generated by an Eulerian–Lagrangian-based LES framework. Before training the machine learning framework, the LES model was validated against measurements for both non-vaporizing and vaporizing sprays. The training points were identified by the momentum flux ratio, which varied between 5 and 120. The Gaussian process-based machine learning algorithm was then trained using the LES data at momentum flux ratios determined based on a design of experiments study. To test the accuracy of the surrogate model, emulated results at representative test points of q = 7 and 40, were compared against LES data in the mid-plane of the geometry. The model predicted spatial variations of the velocity magnitude, pressure, temperature, vapor fraction, TKE, liquid penetration, and SMD with reasonable accuracy (varying between 0.05 and 13.5%). Corresponding to these emulated results, a speedup of 65 was achieved. Based on these observations, we hope that this methodology can be leveraged (and further improved) for conducting large scale parameter sweeps and uncertainty quantification studies for the design of devices based on turbulent multiphase flows.

The algorithm and methodology developed in this study can be applied to other fields of science and engineering for steady-state or statistically stationary flows. Except for selecting a kernel function to capture the signal variation across training datasets, the methodology remains unchanged. For spatiotemporally evolving flows, it is recommended to reduce the dimensionality, especially in the time domain, using some kind of modal decomposition technique for computational efficiency and improvements in speedup. Application of the algorithm used in this paper to spatiotemporally varying flows has been successfully demonstrated in our previous work [17]. While this paper represents the first effort in predicting turbulent multiphase flows using machine learning techniques, it is just the first

step towards the development of a comprehensive framework that can be used for engineering design. The next step is to develop machine learning algorithms capable of predicting fluid dynamic behaviors when multiple independent parameters are varied simultaneously, which is a closer representation of the design space that is encountered in the preliminary and detailed design stages.

Author Contributions: Conceptualization, P.K.; methodology, P.K.; software, H.G., M.K., and P.K.; validation, H.G. and M.K.; formal analysis, H.G. and M.K.; investigation, H.G. and M.K.; resources, P.K.; data curation, H.G. and M.K.; writing—original draft preparation, H.G. and M.K.; writing—review and editing, P.K.; visualization, H.G., M.K., and P.K.; supervision, P.K.; project administration, P.K.; funding acquisition, P.K. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, H.-G.; Khare, P.; Sung, H.-G.; Yang, V. A Large-Eddy-Simulation study of combustion dynamics of Bluff-Body stabilized flames. *Combust. Sci. Technol.* **2016**, *188*, 924–952. [CrossRef]
2. Yoo, Y.-L.; Han, D.-H.; Hong, J.-S.; Sung, H.-G. A large eddy simulation of the breakup and atomization of a liquid jet into a cross turbulent flow at various spray conditions. *Int. J. Heat Mass Transf.* **2017**, *112*, 97–112. [CrossRef]
3. Jiang, X.; Siamas, G.A.; Jagus, K.; Karayiannis, T.G. Physical modelling and advanced simulations of gas-liquid two-phase jet flows in atomization and sprays. *Prog. Energy Combust. Sci.* **2010**, *36*, 131–167. [CrossRef]
4. Moin, P.; Apte, S.V. Large-eddy simulation of realistic gas turbine combustors. *AIAA J.* **2006**, *44*, 698–708. [CrossRef]
5. Madabhushi, R.K. A Model for Numerical Simulation of Breakup of a Liquid Jet in Crossflow. *At. Sprays* **2003**, *13*, 413–424. [CrossRef]
6. Erlebacher, G.; Hussaini, M.Y.; Speziale, C.G.; Zang, T.A. Toward the large-eddy simulations of compressible turbulent flows. *J. Fluid Mech.* **1992**, *238*, 155–185. [CrossRef]
7. Broumand, M.; Birouk, M. Liquid jet in a subsonic gaseous crossflow: Recent progress and remaining challenges. *Prog. Energy Combust. Sci.* **2016**, *57*, 1–29. [CrossRef]
8. Song, J.; Cain, C.C.; Lee, J.G. Liquid Jets in Subsonic Air Crossflow at Elevated Pressure. *J. Eng. Gas Turbines Power* **2014**, *137*. [CrossRef]
9. Lee, K.; Aalburg, C.; Diez, F.J.; Faeth, G.M.; Sallam, K.A. Primary Breakup of Turbulent Round Liquid Jets in Uniform Crossflows. *AIAA J.* **2007**, *45*, 1907–1916. [CrossRef]
10. Vich, G.; Ledoux, M. Investigation of a liquid jet in a subsonic cross-flow. *Int. J. Fluid Mech. Res.* **1997**, *24*, 1–12. [CrossRef]
11. Kihm, K.D.; Lyn, G.M.; Son, S.Y. Atomization of Cross-Injecting Sprays Into Convective Air Stream. *At. Sprays* **1995**, *5*, 417–433. [CrossRef]
12. Zang, Y.; Street, R.L.; Koseff, J.R. A dynamic mixed subgrid—Scale model and its application to turbulent recirculating flows. *Phys. Fluids A Fluid Dyn.* **1993**, *5*, 3186–3196. [CrossRef]
13. Birouk, M.; Lekic, N. Liquid Jet Breakup in Quiescent Atmosphere: A Review. *At. Sprays* **2009**, *19*, 501–528. [CrossRef]
14. Sallam, K.A.; Dai, Z.; Faeth, G.M. Liquid breakup at the surface of turbulent round liquid jets in still gases. *Int. J. Multiph. Flow* **2002**, *28*, 427–449. [CrossRef]
15. Mazallon, J.; Dai, Z.; Faeth, G.M. Primary breakup of nonturbulent round liquid jets in gas crossflows. *At. Sprays* **1999**, *9*, 291–311. [CrossRef]
16. Wu, P.-K.; Kirkendall, K.A.; Fuller, R.P.; Nejad, A.S. Breakup Processes of Liquid Jets in Subsonic Crossflows. *J. Propuls. Power* **1997**, *13*, 64–73. [CrossRef]

17.　Ganti, H.; Khare, P. Data-Driven Surrogate Modeling of Multiphase Flows Using Machine Learning Techniques. *Comput. Fluids* **2020**, *211*, 104626. [CrossRef]

18.　Wu, P.-K.; Faeth, G.M. Aerodynamic Effects on Primary Breakup of Turbulent Liquids. *At. Sprays* **1993**, *3*, 265–289. [CrossRef]

19.　Amighi, A.; Ashgriz, N. Global Droplet Size in Liquid Jet in a High-Temperature and High-Pressure Crossflow. *AIAA J.* **2019**, *57*, 1260–1274. [CrossRef]

20.　Oh, J.; Lee, J.G.; Lee, W. Vaporization of a liquid hexanes jet in cross flow. *Int. J. Multiph. Flow* **2013**, *57*, 151–158. [CrossRef]

21.　Stenzler, J.N.; Lee, J.G.; Santavicca, D.A.; Lee, W. Penetration of liquid jets in a cross-flow. *At. Sprays* **2006**, *16*, 887–906. [CrossRef]

22.　Rachner, M.; Julian, B.; Christoph, H.; Thomas, D. Modelling of the atomization of a plain liquid fuel jet in crossflow at gas turbine conditions. *Aerosp. Sci. Technol.* **2002**, *6*, 495–506. [CrossRef]

23.　Wang, Q.; Hesthaven, J.S.; Ray, D. Non-intrusive reduced order modeling of unsteady flows using artificial neural networks with application to a combustion problem. *J. Comput. Phys.* **2019**, *384*, 289–307. [CrossRef]

24.　Takbiri-Borujeni, A.; Kazemi, H.; Nasrabadi, N. A data-driven proxy to Stoke's flow in porous media. *arXiv* **2019**, arXiv:1905.06327.

25.　Yeh, S.-T.; Wang, X.; Sung, C.-L.; Mak, S.; Chang, Y.-H.; Zhang, L.; Wu, C.F.J.; Yang, V. Common Proper Orthogonal Decomposition-Based Spatiotemporal Emulator for Design Exploration. *AIAA J.* **2018**, *56*, 2429–2442. [CrossRef]

26.　Moiz, A.A.; Pal, P.; Probst, D.; Pei, Y.; Zhang, Y.; Som, S.; Kodavasal, J.A. A Machine Learning—Genetic Algorithm (ML-GA) Approach for Rapid Optimization Using High-Performance Computing. *SAE Int. J. Commer. Veh.* **2018**, *11*, 291–306. [CrossRef]

27.　Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations. *arXiv* **2017**, arXiv:1711.10561.

28.　Bai, Z.; Brunton, S.L.; Brunton, B.W.; Kutz, J.N.; Kaiser, E.; Spohn, A.; Noack, B.R. Data-Driven Methods in Fluid Dynamics: Sparse Classification from Experimental Data. In *Whither Turbulence and Big Data in the 21st Century*; Pollard, A., Castillo, L., Danaila, L., Glauser, M., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 323–342.

29.　Brunton, S.; Proctor, J.L.; Kutz, J.N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3932–3937. [CrossRef]

30.　Moonen, P.; Allegrini, J. Employing statistical model emulation as a surrogate for CFD. *Environ. Model. Softw.* **2015**, *72*, 77–91. [CrossRef]

31.　Bright, I.; Lin, G.; Kutz, J.N. Compressive sensing based machine learning strategy for characterizing the flow around a cylinder with limited pressure measurements. *Phys. Fluids* **2013**, *25*. [CrossRef]

32.　Kamin, M.; Khare, P. Liquid jet in crossflow: Effect of momentum flux ratio on spray and vaporization characteristics. In Proceedings of the ASME Turbo Expo 2019: Turbomachinery Technical Conference and Exposition, Phoenix, AZ, USA, 17–21 June 2019.

33.　Khare, P.; Wang, S.; Yang, V. Modeling of finite-size droplets and particles in multiphase flows. *Chin. J. Aeronaut.* **2015**, *28*, 974–982. [CrossRef]

34.　Hadfield, S.; Lawrence, N.D.; Wilkinson, R.; Durrande, N. Gaussian Process Summer School. 2017. Available online: http://gpss.cc/gpss17/ (accessed on 2 September 2020).

35.　Ebden, M. Gaussian processes: A quick introduction. *arXiv* **2015**, arXiv:1505.02965.

36.　Rasmussen, C.E. *Gaussian Processes in Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 63–71. [CrossRef]

37.　Chen, Z.; Wang, B. How priors of initial hyperparameters affect Gaussian process regression models. *Neurocomputing* **2018**, *275*, 1702–1710. [CrossRef]

38.　GPSS. GPy: A Gaussian Process Framework in Python. 2014. Available online: http://github.com/SheffieldML/GPy (accessed on 2 September 2020).

39.　Viana, F.A.C. A Tutorial on Latin Hypercube Design of Experiments. *Qual. Reliab. Eng. Int.* **2016**, *32*, 1975–1985. [CrossRef]

40.　Shields, M.D.; Zhang, J. The generalization of Latin hypercube sampling. *Reliab. Eng. Syst. Saf.* **2016**, *148*, 96–108. [CrossRef]

41. Oefelein, J.C.; Yang, V. Modeling High-Pressure Mixing and Combustion Processes in Liquid Rocket Engines. *J. Propuls. Power* **1998**, *14*, 843–857. [CrossRef]

42. Reitz, R. Modeling atomization processes in high-pressure vaporizing sprays. *At. Spray Technol.* **1987**, *3*, 309–337.

43. Liu, A.B.; Mather, D.; Reitz, R.D. Modeling the effects of drop drag and breakup on fuel sprays. *SAE Trans.* **1993**, *102*, 83–95.

44. O'Rourke, P.J.; Amsden, A.A. The TAB Method for Numerical Calculation of Spray Droplet Breakup. SAE Technical Paper 872089. 1987. Available online: https://www.sae.org/publications/technical-papers/content/872089/ (accessed on 10 August 2020). [CrossRef]

45. Mundo, C.; Sommerfeld, M.; Tropea, C. Droplet-wall collisions: Experimental studies of the deformation and breakup process. *Int. J. Multiph. Flow* **1995**, *21*, 151–173. [CrossRef]